

Two New Rapid SNP-Typing Methods for Classifying *Mycobacterium tuberculosis* Complex into the Main Phylogenetic Lineages

David Stucki^{1,2}, Bijaya Malla^{1,2}, Simon Hostettler^{1,2}, Thembela Huna³, Julia Feldmann^{1,2}, Dorothy Yeboah-Manu⁴, Sonia Borrell^{1,2}, Lukas Fenner⁵, Iñaki Comas^{6,7}, Mireia Coscollà^{1,2}, Sebastien Gagneux^{1,2*}

1 Swiss Tropical and Public Health Institute, Basel, Switzerland, **2** University of Basel, Basel, Switzerland, **3** Division of Mycobacterial Research, Medical Research Council, National Institute for Medical Research, London, United Kingdom, **4** Noguchi Memorial Institute for Medical Research, University of Ghana, Legon, Ghana, **5** Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland, **6** Genomics and Health Unit, Centre for Public Health Research, Valencia, Spain, **7** CIBER Epidemiología y Salud Pública, Madrid, Spain

Abstract

There is increasing evidence that strain variation in *Mycobacterium tuberculosis* complex (MTBC) might influence the outcome of tuberculosis infection and disease. To assess genotype-phenotype associations, phylogenetically robust molecular markers and appropriate genotyping tools are required. Most current genotyping methods for MTBC are based on mobile or repetitive DNA elements. Because these elements are prone to convergent evolution, the corresponding genotyping techniques are suboptimal for phylogenetic studies and strain classification. By contrast, single nucleotide polymorphisms (SNP) are ideal markers for classifying MTBC into phylogenetic lineages, as they exhibit very low degrees of homoplasmy. In this study, we developed two complementary SNP-based genotyping methods to classify strains into the six main human-associated lineages of MTBC, the “Beijing” sublineage, and the clade comprising *Mycobacterium bovis* and *Mycobacterium caprae*. Phylogenetically informative SNPs were obtained from 22 MTBC whole-genome sequences. The first assay, referred to as MOL-PCR, is a ligation-dependent PCR with signal detection by fluorescent microspheres and a Luminex flow cytometer, which simultaneously interrogates eight SNPs. The second assay is based on six individual TaqMan real-time PCR assays for singleplex SNP-typing. We compared MOL-PCR and TaqMan results in two panels of clinical MTBC isolates. Both methods agreed fully when assigning 36 well-characterized strains into the main phylogenetic lineages. The sensitivity in allele-calling was 98.6% and 98.8% for MOL-PCR and TaqMan, respectively. Typing of an additional panel of 78 unknown clinical isolates revealed 99.2% and 100% sensitivity in allele-calling, respectively, and 100% agreement in lineage assignment between both methods. While MOL-PCR and TaqMan are both highly sensitive and specific, MOL-PCR is ideal for classification of isolates with no previous information, whereas TaqMan is faster for confirmation. Furthermore, both methods are rapid, flexible and comparably inexpensive.

Citation: Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, et al. (2012) Two New Rapid SNP-Typing Methods for Classifying *Mycobacterium tuberculosis* Complex into the Main Phylogenetic Lineages. PLoS ONE 7(7): e41253. doi:10.1371/journal.pone.0041253

Editor: Igor Mokrousov, St. Petersburg Pasteur Institute, Russian Federation

Received: April 20, 2012; **Accepted:** June 19, 2012; **Published:** July 20, 2012

Copyright: © 2012 Stucki et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Swiss National Science Foundation (grant number PP00A-119205), the National Institutes of Health (AI090928 and HHSN266200700022C), and the Leverhulme-Royal Society Africa Award (AA080019). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Corresponding author SG is a PLoS ONE Editorial Board member. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: Sebastien.Gagneux@unibas.ch

Introduction

Genotyping of human-associated *Mycobacterium tuberculosis* complex (MTBC) plays an increasing role for understanding the epidemiology and biology of tuberculosis (TB) [1,2]. On the one hand, genotyping techniques are key for standard molecular epidemiological investigations of TB such as defining chains of ongoing transmission and differentiating patient relapse from exogenous re-infection [1]. On the other hand, strain genotyping in MTBC is important for defining the evolutionary background of clinical isolates. There is mounting evidence that the strain genetic background might influence the outcome of TB infection and disease. Experimental studies have shown that clinical strains differ in immunogenicity and virulence, but whether this variation

translates into comparable clinical phenotypes is unclear [2–9]. To study the effect of strain variation and detect relevant genotype-phenotype associations, suitable phylogenetic markers and appropriate genotyping methods are required [10].

Several methods for genotyping of MTBC have been developed over the past years [11]. Two of the most popular methods, spoligotyping [12] and MIRU-VNTR [13], rely on repetitive DNA elements, and are currently considered the gold standard for TB transmission studies [1,13]. However, due to their propensity for convergent evolution and resulting homoplasies [10], inferring robust phylogenies using these methods can be problematic. Moreover, because of these homoplasies, classification of MTBC strains into robust phylogenetic lineages is not always possible and

misclassification can occur [14,15]. In addition to mobile genetic elements (e.g. IS6110 [16]), repetitive DNA (DR region [17], MIRU-VNTR [18]), and large sequence polymorphisms [19,20], single nucleotide polymorphisms (SNPs) have become available as phylogenetic markers for MTBC. SNPs are ideal markers for genotyping of MTBC, as they represent unique events and show virtually no homoplasy [10]. Several studies have reported the use of SNPs for phylogenetic classification of MTBC [21–29]. However, no consensus has yet been reached as to what set of SNPs should be used for standard phylotyping of MTBC. Importantly, not all phylogenetically informative SNPs will be equally amenable to all possible SNP-typing platforms [30,31]. In MTBC, some of the methods that have been proposed are allele-specific PCR [32], real-time PCR [33], Sanger sequencing [34], SNaPshot minisequencing [28], iPLEX Gold (SEQUENOM Inc.) [29] and MLPA [25]. All of these methods differ in their technical requirements. Hence, depending on the particular technique used, different, but phylogenetically equivalent SNP sets will be required. In addition, these typing methods vary largely in their throughput, cost, and flexibility. With respect to the latter, we identified a lack of methods, which are at the same time flexible and rapid (<1 day/96 samples), and amenable to a variable number of strains and a limited number of SNPs (5–50).

Recently, whole genome sequences of a global collection of 21 MTBC strains representing the six main lineages of human-associated MTBC became available [35]. The 9,037 SNPs identified in these genomes represent an ideal starting point for extracting diagnostic or “canonical” SNPs, which are lineage-specific, to design phylogenetically robust genotyping assays [36]. In this study, we present two new and complementary SNP-typing methods for MTBC. These methods are based on two phylogenetically equivalent sets of SNP markers that are specific for the 6 main human-associated lineages of MTBC. Additionally, we present SNPs specific for the clade comprising *Mycobacterium bovis* and *Mycobacterium caprae* [26], and for the “Beijing”-sublineage of Lineage 2 (East-Asian) (Figure 1). “Beijing” strains are of special epidemiological interest because of their repeated association with drug resistance and hypervirulence in infection models [37,38]. The *M. bovis*/*M. caprae* specific SNP does not capture the other animal-associated strains, i.e. *M. pinnipedii*, *M. microti*, *M. orygis*, *M. mungi* and the Dassie bacillus.

The first SNP-typing method is MOL-PCR (multiplexed oligonucleotide ligation PCR) [39], which uses allele-specific ligation for allele discrimination, singleplex PCR for signal amplification and fluorescent microspheres (beads) for the signal detection on a flow cytometer (Figure 2, see Methods). A Luminex device is needed; i.e. a flow cytometric platform for various nucleic acid and immunological assays [40–42], which allows for simultaneous interrogation of 50 biallelic SNPs. MOL-PCR is flexible for both the number of SNPs as well as for the number of strains tested (individual tubes to 96-well plates). We developed an 8-plex SNP-typing assay for the identification of the main phylogenetic lineages of MTBC. As a second method, we present SNP-typing with TaqMan real-time PCR. As a commercially available system, TaqMan has proved to be sensitive and specific in numerous studies for various applications and species [43–46], and is therefore considered the standard for SNP-typing in our laboratory [14,20]. Reactions are performed in a 96-well format in less than 2 hours. In contrast to MOL-PCR, only one SNP is assessed per reaction.

Methods

Ethics statement

All MTBC strains used in this study were from published reference collections [14,20,26,47–49] or from an ongoing molecular epidemiology study on tuberculosis in Nepal. This study was ethically approved by the Nepal Health Research Council (NHRC), Kathmandu, Nepal, and the Ethics Committee of Basel, Switzerland (EKBB). Written informed consent was obtained from all Nepalese patients.

Strain panels

Two different strain panels were used in this study. A “training panel” of 46 MTBC strains was first chosen for establishing the MOL-PCR and TaqMan assays. These strains had been characterized previously by spoligotyping and were available as crude lysates and purified DNA. This training panel contained clinical strains from published reference collections [14,20,26,47–49] and some clinical isolates from Nepal (manuscript in preparation). The training panel was composed of five to seven isolates of the 6 main phylogenetic lineage of MTBC, and additionally, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium pinnipedii* and *Mycobacterium microti* as controls for the *M. bovis*/*M. caprae*-specific SNP in the MOL-PCR assay. For each of these strains, SNP-typing was done by MOL-PCR and TaqMan for both crude lysates and purified DNA. Allele calls, i.e. lineage assignments, were obtained for all data points, and results between the two methods compared. As it was not possible to use the same SNPs for both MOL-PCR and TaqMan (due to probe design restrictions), we compared the lineage assignment rather than the locus-specific alleles.

A “test panel” was then chosen for the validation of MOL-PCR and TaqMan under “blind” conditions. This test panel consisted of 78 clinical MTBC isolates recently obtained from a molecular epidemiology study in Nepal. No previous genotyping had been done for these samples. The samples were cultured on Löwenstein-Jensen (LJ) medium, frozen in tryptic soy agar and glycerol, and heat-inactivated for 1 h at 90°C. Purified DNA was available for fifteen of these strains.

Culturing and DNA extraction

MTBC strains were grown in standard Middlebrook 7H9 medium supplemented with ADC (Becton Dickinson, Allschwil, Switzerland), Tween80 (Sigma-Aldrich) and glycerol (Sigma-Aldrich). Liquid medium and agar plates for culturing *Mycobacterium africanum* (i.e. MTBC Lineage 5 and 6) were supplemented with 40 mM sodium pyruvate (Sigma-Aldrich). All bacterial cultures were incubated at 37°C. After two weeks, 500 µL culture were heat-inactivated (1 hour at 90°C) and centrifuged. The supernatant was used as input for TaqMan and MOL-PCR. Alternatively, 100 µL of a frozen stock culture was thawed and heat-inactivated (100 µL culture added to 100 µL TE buffer, 1 hour at 90°C). From all cultures, purified genomic DNA was extracted with the CTAB method [50]. DNA concentration was determined with a Nanodrop ND-1000 device (Thermo Scientific NanoDrop Products, Wilmington, USA).

Whole genome sequencing and SNP identification

Informative SNPs were obtained from whole genome sequencing as described before [35,51]. In brief, short reads from Illumina-sequencing of 22 MTBC strains were mapped to the reference genome of H37Rv and strain-specific nucleotide differences were extracted as SNPs. A neighbor-joining phylogenetic tree was constructed with MEGA [52], and SNPs mapped to

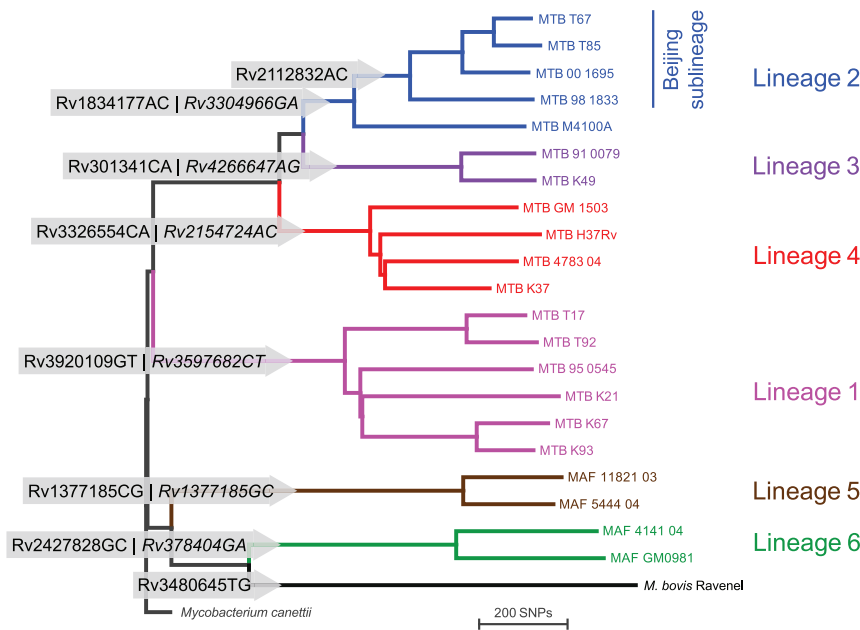


Figure 1. Phylogenetic tree of 22 whole genome sequences of MTBC plus *Mycobacterium canettii* as outgroup, and canonical SNPs used for MOL-PCR and TaqMan assays. Rv-numbers represent SNPs used for MOL-PCR and TaqMan in this study, respectively, and nucleotide change at the position in the annotated reference sequence of H37Rv [56]. The first number indicates the SNP used for MOL-PCR, and the italic number the SNP used for TaqMan. Assays for the “Beijing” sublineage of Lineage 2 and for *M. bovis*/*M. caprae* were developed only for MOL-PCR (modified from [51]).

doi:10.1371/journal.pone.0041253.g001

the tree using Mesquite [53]. Clade-specific SNPs were compiled in OpenOffice spreadsheets (openoffice.org). Position in reference to H37Rv, codon change, essentiality [54,55], and annotated function of the gene [56] were collected for each SNP. This list was used as a starting point for oligonucleotide design. In the course of the study, 151 additional genome sequences became available (manuscript in preparation). With the total of 172 genome sequences representing the global MTBC diversity, we confirmed that the SNPs chosen were specific for all members of a lineage. In reference to the first MTBC genome sequenced and annotated [56] (Genbank AL123456.2), every SNP was named with the prefix “Rv”, followed by 7 digits that represent the position of this SNP in the genome of H37Rv, and the nucleotide change (i.e. “Rv2154724AC” for the Lineage 4-specific SNP in *katG*, codon 463).

Assay description Luminex MOL-PCR

The assay described here is conceptually related to Multiplex Ligation-dependent Probe Amplification (MLPA) that was also described for SNP-typing of *M. tuberculosis* [25,57]. MLPA uses allele-specific ligation and stuffer sequences of variable lengths, which allow multiplexed typing with a capillary electrophoresis device. The high level of multiplexing is made possible by the PCR-amplification of the ligated oligonucleotides rather than the amplification of template DNA. This allows the use of a universal primer-pair for the singleplex-amplification of all ligated oligonucleotides and avoids multiplex-PCR. Allele-specific ligation was recently combined with the Luminex platform and named MOL-PCR, and used for pathogen detection and SNP-typing of *Bacillus anthracis*, *Yersinia pestis*, and *Francisella tularensis* [39,58]. MOL-PCR uses fluorescent beads and coupled *tag*-sequences instead of stuffer sequences for the multiplexed analysis. This allows the signal detection on a flow cytometer, i.e. a Luminex device. We adapted and modified MOL-PCR for typing of an 8-plex set of MTBC

lineage-specific SNPs. In our assay, the *tag*-sequences complementary to *antiTag*-sequences on microspheres were included on the left-probe-oligonucleotide (LPO), i.e. preceding the 5'-end of the allele-discriminating sequence (left-hybridizing sequence, LHS), contrary to the original publication of MOL-PCR [39], where *tag*-sequences were included on the right-probe-oligonucleotide (RPO). With this modification in oligonucleotide design, two competing LPO could be used, differing only by the 3'-end (the SNP of interest) and the *tag*-sequence. This change allowed the calculation of an allelic ratio rather than a signal-to-noise ratio, and improved the sensitivity by discriminating background signal from uncalled allele signal. An additional change was made in separating the hybridization/ligation step from the PCR step. We obtained high background signal levels when combining these two steps. The assay therefore consisted of three steps (Figure 2): 1. Competitive hybridization of allele-specific probes and ligation to universal, 3'-adjacent probes. This step provides the allele-specificity of the assay. 2. Signal amplification by PCR and a fluorescently labeled primer, which guarantees the sensitivity. 3. Detection of allele-specific signals by hybridizing amplicons to allele-specific beads and read-out by flow cytometry.

SNP selection and MOL-PCR oligonucleotide design

The selection of suitable SNPs is crucial for the success of a typing assay. In MOL-PCR, probes of 20–45 bp are designed to anneal directly adjacent to the SNP of interest. The need for a suitable sequence up- and downstream of the SNP requires a number of SNPs to choose from as a starting point. Eligible SNPs for MOL-PCR were chosen using following criteria:

1. SNP is specific for the lineage of interest
2. SNP is not in a region where a genomic deletion has been described (to ensure binding of oligonucleotides and subsequent allele-calls for all strains)

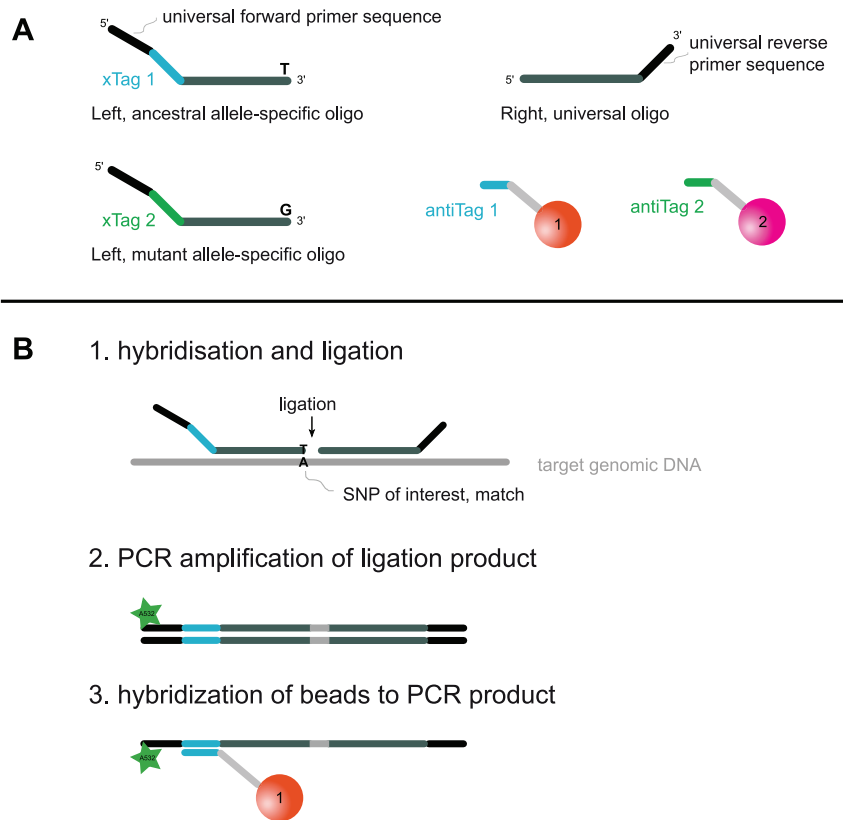


Figure 2. Schematic illustration of MOL-PCR. A. Three oligonucleotides and two fluorescently labelled beads are used for interrogation of one SNP. B. MOL-PCR consists of three steps. 1. One of two competitive allele-specific left-hand probe oligonucleotides (LPO) and one universal right-hand probe oligonucleotide (RPO) are hybridized to the template DNA and ligated. 2. With PCR and a reporter-labelled forward primer, the ligated oligonucleotides (LPO+RPO) are amplified. 3. After denaturation of the PCR product, allele-specific fluorescent beads carrying an allele-specific *antiTag* sequence are hybridized to the amplicons. This will result in beads carrying reporter fluorescence (bead 1 in example) and beads not carrying fluorescence (bead 2). Reporter fluorescence (Alexa532) per bead is measured with a flow cytometric device (Luminex). For the 8-plex assay, a total of 24 oligonucleotides and 16 beads are used in the same reaction.
doi:10.1371/journal.pone.0041253.g002

3. No other SNP less than 50 bp up- or downstream (to avoid mismatches in oligonucleotide annealing)
4. SNP is not a transition (i.e. no nucleotide change A>G, G>A, C>T or T>C; lower unspecific signal because no alternate base pairing possible)
5. SNP is in a coding sequence (deletions or insertions less likely)
6. SNP is in an essential gene (deletions less likely)
7. SNP change is synonymous (no amino acid change; therefore lower selective pressure acting on SNP compared to non-synonymous SNP)

For each eligible SNP, 100 bp of enclosing genomic sequence of H37Rv up- and downstream of the SNP were obtained using Eugene [59] and saved as multifasta file. Sequences and SNPs were introduced in AlleleID (Premier Biosoft, Palo Alto, USA) and suitable oligonucleotides searched using the MLPA module. As MLPA uses stuffer sequences rather than *tag*-sequences (for hybridizing to microspheres with coupled *antiTag*-oligonucleotides), the latter were introduced manually and according to sequences provided by Luminex Corporation (Austin, USA). We used the following criteria for oligonucleotide search in AlleleID: 1. GC-content between 35% and 65%, 2. T_m between 72°C and 90°C, 3. Minimum left and right hybridization sequence (LHS and RHS) length 21 bp. Probes with high scores in AlleleID were exported from AlleleID to OpenOffice spreadsheets. BLAST search against

M. tuberculosis H37Rv was run to exclude cross-homology. The 24 oligonucleotides for each lineage-specific SNP were pooled (3 oligonucleotides each) and assessed for heterodimer formation with SBEprimer [60], AutoDimer [61] and Oligoanalyzer (www.idtdna.com). Probes with potential dimer formation (dG<-11 kcal/mol) were excluded, a new SNP chosen and new probes designed, until a compatible set was found. The final set of 24 oligonucleotides was synthesized by Sigma-Aldrich (Buchs, Switzerland). The 5'-end of each RPO was phosphorylated to enable ligation.

Luminex MOL-PCR procedure

All SNPs were interrogated in one tube (multiplexed). Our modified MOL-PCR assay consisted of three laboratory steps. 1. Oligonucleotides were hybridized to target DNA and ligated. 2. The ligation product was amplified by singleplex PCR (one primer pair for all SNPs). 3. Beads were hybridized to the amplification products and run on a flow cytometric device.

Hybridization and ligation. In a 10 µL volume with H₂O, 0.5 µL of oligonucleotide-mix (40 nM each oligonucleotide; see Table 1) were mixed with 1 µL of ligation buffer (NEB; New England Biolabs, Ipswich, USA), 0.1 µL of thermostable ligase (NEB) and 3 µL of pure DNA (10 ng/µL) or 4 µL heat-inactivated, crude extract. In a thermocycler, the following

Table 1. Oligonucleotides used for Luminex MOL-PCR.

MTBC Lineage ¹	SNP ²	Gene	Ess/Syn ³	Strand ⁴	Allele ⁵	Bead ⁶	Sequence ⁷
1	Rv3920109GT	yrbE4A	ess/syn	antisense	LPO_ancestral	30	GGGTTCCCTAAGGGTTGGACTtaacatttaactctataacacGTAGTCATGACCGCTATAGGTGTGCGAATC
					LPO_mutant	35	GGGTTCCCTAAGGGTTGGACatcttcaataactcttattGTAGTCATGACCGCTATAGGTGTGCGAATA
					RPO		P-TCGGCGCAGTAACAAACGACGGTGATCTAGATTGGATCTTGCTGGCAC
2	Rv1834177AC	rpsA	ess/syn	sense	LPO_ancestral	12	GGGTTCCCTAAGGGTTGGACataatcaatttcaactttcttCGGAGTTCCTGAATAACTTGCAAAAGGCCACCATCCGA
					LPO_mutant	13	GGGTTCCCTAAGGGTTGGACaaatacaataacttcaattcaactCGGAGTTCCTGAATAACTTGCAAAAGGCCACCATCCGC
					RPO		P-AAGGGTGTGTCCTCGATCGTCAACTTCGGTCTAGATTGGATCTTGCTGGCAC
3	Rv301341CA	Rv0249c	ess/syn	sense	LPO_ancestral	18	GGGTTCCCTAAGGGTTGGAacatttcttcaattcaattacTGAGCCGGAATCCAGCAGGAAC
					LPO_mutant	19	GGGTTCCCTAAGGGTTGGAatactttacaacaataacacacTGAGCCGGAATCCAGCAGGAATA
					RPO		P-GGTAATACCATCGCTCCACCGGAATCTCTAGATTGGATCTTGCTGGCAC
4	Rv3326554CA	Rv2971	ess/nonsyn.	antisense	LPO_ancestral	20	GGGTTCCCTAAGGGTTGGactttctctacttctcaactaattGACGAATGTGAGGTGCGATAAGGTTTCGATGTG
					LPO_mutant	21	GGGTTCCCTAAGGGTTGGAtcaactctcaattcttacttaattGACGAATGTGAGGTGCGATAAGGTTTCGATGTT
					RPO		P-CTCCGCGGTGAAGTTGGACACGCCGATCTCTAGATTGGATCTTGCTGGCAC
5	Rv1377185CG	Rv1234	noness/syn	sense	LPO_ancestral	36	GGGTTCCCTAAGGGTTGGAataaacaacttcaactacacaaGCAGGTGACCATCGTTGGCGTGGACCTC
					LPO_mutant	37	GGGTTCCCTAAGGGTTGGAatacaacatcttaataacatacacaGCAGGTGACCATCGTTGGCGTGGACCTG
					RPO		P-ATGCAGGTGAACGGGTCAAGGGCTCTAGATTGGATCTTGCTGGCAC
6	Rv2427828GC	Rv2164c	ess/syn	sense	LPO_ancestral	33	GGGTTCCCTAAGGGTTGGAactatttcaacttaataaCTCGGTGGACAAACACAGTGTGAGG
					LPO_mutant	34	GGGTTCCCTAAGGGTTGGAactatttctactactatacaCATCGGTGGACAAACACAGTGTGAGC
					RPO		P-CCTAGTCCGACCCGACGACCGAATCTAGATTGGATCTTGCTGGCAC
<i>M. bovis</i> / <i>M. caprae</i>	Rv3480645TG	Rv3113	ess/nonsyn	sense	LPO_ancestral	28	GGGTTCCCTAAGGGTTGGAacttaattcttaataatctatcTTCCGCTTGGTACCCGGAGCTTCACT
					LPO_mutant	29	GGGTTCCCTAAGGGTTGGAacttcttataactcaactaaTTCCGCTTGGTACCCGGAGCTTCAACG
					RPO		P-GCAGCCACCCGACCGAGTCATGCTCAAAATCTAGATTGGATCTTGCTGGCAC
Beijing sublineage	Rv2112832AC	Rv1865c	noness/syn	sense	LPO_ancestral	14	GGGTTCCCTAAGGGTTGGAaatttcttcttcttcaaatTTTCGCCATCGCCTCATCGATGTCGCCGATA
					LPO_mutant	25	GGGTTCCCTAAGGGTTGGAactttcttaataacataacacTTCGCCATCGCCTCATCGATGTCGCCGATC
					RPO		P-GCAGCTTGGCTCCCAAGTTGGATCTAGATTGGATCTTGCTGGCAC

¹Nomenclature according to [2].

²Position of SNP in reference to H37Rv.

³SNP in an essential gene ("ess"); synonymous SNP ("syn").

⁴Strand orientation of oligonucleotides.

⁵LPO, left probe oligonucleotide, annealing upstream of SNP of interest and including it; RPO, right probe oligonucleotide, annealing downstream of SNP of interest.

⁶xTag bead with coupled antiTag-sequence.

⁷Universal primer sequences (upper case, not underlined), allele-specific tag-sequences (lower case, not underlined), and sequences hybridizing to the template genomic sequence (underlined). RPO are 5'-phosphorylated.

doi:10.1371/journal.pone.0041253.t001

protocol was run: 4 min 94°C, 30 cycles 25 sec 94°C and 30 sec 50°C.

PCR: In a 10 µL volume, 3 µL of the ligation product was added to the PCR mix of 5.095 µL of H₂O, 1 µL PCR buffer without MgCl₂ (Roche, Basel, Switzerland), 0.2 µL dNTPs (10 mM, Sigma-Aldrich), 0.5 µL forward primer (10 µM; Sigma-Aldrich, 5'-Alexa532-GGGTTCCTAAGGGTTGGA), 0.125 µL reverse primer (10 µM; Sigma-Aldrich, GTGCCAG-CAAGATCCAATCTAGA) and 0.08 µL FastStart Taq polymerase (Roche). The following protocol was run in a thermocycler: 2 min 96°C and 43 cycles of 20 sec 94°C, 20 sec 58°C, 20 sec 72°C.

Hybridization of beads and signal detection. A bead mix containing 2 µL of beads (16 polystyrene bead regions of MagPlex-Tag beads, Luminex Corp., Austin, USA; Table 1) and 3 µL of buffer (1.33 M NaCl (Merck Chemicals, Darmstadt, Germany), 83.33 mM MES (Sigma-Aldrich)) was added to the 10 µL PCR product. In a thermocycler, the following protocol was run to hybridize beads (*antiTags*) to PCR products (*tags*): 1 min 94°C, ramp-down 0.1°C/sec decrement to 25°C, 5 min 25°C. 80 µL of running buffer (10 mM Tris-HCl, 0.1 mM EDTA (Sigma-Aldrich), 90 mM NaCl, 0.04% (v/v) Triton X-100 (Sigma-Aldrich), pH 8) were added and the reaction transferred to flat-bottom plates (BioRad, Hercules, USA). A BioPlex 200 device (BioRad) was used to discriminate beads and measure reporter fluorescence (Alexa532) for each bead, i.e. each allele. Results were obtained as reporter median fluorescence intensity (MFI) per bead region and sample, and exported from BioPlex Manager software into spreadsheets.

Allele calling MOL-PCR

For each sample, the allelic state of each SNP was assessed with the following algorithm:

1. Bead counts >40 beads per bead region
2. Allelic ratio (AR)<0.4; ancestral allele called
3. Allelic ratio (AR)>0.6; mutant allele called

$$AR = (MFI_{\text{mutant}}/MFI_{H_2O_mutant}) / ((MFI_{\text{mutant}}/MFI_{H_2O_mutant}) + (MFI_{\text{ancestral}}/MFI_{H_2O_ancestral}))$$

4. MFI>threshold MFI (see below)

MFI = median fluorescence intensity of reporter-dye (Alexa532) of the corresponding allele/bead region

A threshold MFI for each allele/bead was used to exclude low signal calls and to avoid SNP-calls for H₂O signals (false-positives). This value was different for each SNP and each allele (i.e. each bead region) and had to be calculated for each run, as we observed significant differences of H₂O signal levels between different runs, and occasionally H₂O signal levels higher than uncalled allele signals. The pragmatic approach that we used to calculate the signal threshold for allele was the following. For each allele, the threshold was set higher than each of the corresponding H₂O values (a minimum of three H₂O samples per 96-well-plate were included). The threshold was then lowered so far that no allele was called for any of the H₂O-sample, which is made possible by the range of the allelic ratio. Furthermore, an absolute minimal threshold value was used.

Assay description TaqMan PCR

TaqMan real-time PCR was originally described in 1991 [62] and was since then optimized for commercial use. The PCR based method makes use of fluorescently labeled allele-specific probes,

the 5'-3'-exonuclease activity of Taq polymerase and a non-fluorescent quencher. During PCR amplification of the region of interest, fluorescence is detected in a real-time PCR thermocycler.

SNP selection and probe design for TaqMan

Primers and probes for Lineage 2, Lineage 3 and Lineage 4 were described before (Table 2). SNP for Lineage 1 was chosen from the list of lineage-specific SNPs recently published [10], and SNPs for Lineage 5 and 6 from the list of genome sequences as described above. For the design of probes and primers, the software Primer Express (Applied Biosystems, Carlsbad, USA) was used. As TaqMan assays were singleplex, oligonucleotide design was less complex, as heterodimer formation is limited to the two primers and two probes of one assay. Furthermore, the design of the primer pair can be varied over a large sequence range.

TaqMan real-time PCR

TaqMan real-time PCR was performed according to standard protocols (Applied Biosystems, Carlsbad, USA). Briefly, in a 12 µL volume, 2 µL DNA sample were added to a mix of 5 µL TaqMan Universal MasterMix II (Applied Biosystems), 5 µL H₂O containing forward and reverse primer (0.21 µM each; Sigma-Aldrich), probe A for ancestral allele and probe B for mutant allele (0.83 µM each; Applied Biosystems). Primer and probe sequences are listed in Table 2. Reactions were run in a StepOnePlus thermocycler (Applied Biosystems; 60°C 30 sec; 95°C 10 min; 95°C 15 sec and 60°C 1 min for 40 cycles; 60°C 30 sec) and fluorescence intensity in the VIC and FAM channels measured at the end of every cycle. Results were analyzed with StepOne software (Applied Biosystems) and alleles called with the default algorithm.

Spoligotyping

43-spacer spoligotyping was performed following standard protocols on a membrane [12].

PCR and sequencing of putative deletion in Rv3113

As we did not obtain MOL-PCR signal for the *M. bovis*/*M. caprae*-specific SNP for two strains (N1007 and N1032, see Table S1), the locus was assessed for a potential deletion in Rv3113. PCR was run in a 25 µL volume with 12.3 µL H₂O, 2.5 µL PCR buffer with MgCl₂ (Roche), 5 µL GC buffer, 0.75 µL forward primer K-297 (10 µM, Sigma-Aldrich, CCATGATGCTGGCA-GAACTGA), 0.75 µL reverse primer K-298 (10 µM, Sigma-Aldrich, CCTGCGTACCTTCGTCGTCA), 0.5 µL dNTPs (10 mM, Qiagen, Hombrechtikon, Switzerland) and 0.2 µL Fast Start Taq Polymerase (Roche). Reaction was run in a standard thermocycler (96°C, 7 min; 35 cycles of 96°C 30 sec, 62°C 30 sec, 72°C 30 sec; 72°C 4 min). PCR product was analyzed using standard gel electrophoresis and stained with ethidium bromide.

Results

Eight SNPs were used in the multiplexed MOL-PCR assay (Table 1). Six SNPs were interrogated serially with TaqMan assays (Table 2). Forty-six MTBC strains were run as a "training panel" (Table S1). Most samples were available as 10 ng/µL purified genomic DNA (CTAB-method) and as crude lysates from 7H9 culture. Both methods, MOL-PCR and TaqMan, were run in parallel for all samples and all SNPs. Table S1 shows allelic state for each SNP for both purified DNAs and crude extracts, called by MOL-PCR and TaqMan, respectively.

For MOL-PCR, 662 of 672 (98.6%) total data points (alleles) resulted in a successful call of either the ancestral or the mutant

Table 2. Primer and probe sequences for TaqMan assays.

MTBC Lineage ¹	LSP name ²	Spoligotype name ³	SNP ⁴	Forward primer	Reverse primer	Ancestral allele probe ⁵	Mutant allele probe ⁵	Reference
1	Indo-oceanic	EAI, MANU1	Rv3597682CT	TGTCAACGAAGGCGATCAGA	GACCGTTCGGCAGCTT	6FAM-ACAAGGGCGACGTC-MGBNFQ	VIC-ACAAGGGCGACATC-MGBNFQ	This study
2	East Asian	Beijing	Rv3304966GA	CCTTCGATGTTGTGCTCAATGT	CATGCGGCGATCTCATTGT	6FAM-CCCAGGAGGGGTAC-MGBNFQ	VIC-CCCAGGAAAGGTACT-MGBNFQ	[14]
3	East-African-Indian	CAS	Rv4266647AG	GCATGGATGCGTTGAGATGA	CGAGTCGACGCGACATACC	VIC-AAGAATGCAGCTTGTGA-MGBNFQ	6FAM-AAGAATGCAGCTTGTGCA-MGBNFQ	[14]
4	Euro-American	X, Haarlem, LAM, Uganda	Rv2154724AC	CCGAGATTGCCAGCCTTAAG	GAAACTAGCTGTGAGACAGTCAATCC	VIC-CCAGATCCTGGCATC-MGBNFQ	6FAM-CAGATCCGGGCATC-MGBNFQ	[20]
5	<i>M. africanum</i> West African I	AFRI 2	Rv1377185GC	TCCAGCAGGTGACCATCGT	GGCTGTGACCGCTTCAAC	VIC-CGTGGACCTCATG-MGBNFQ	6FAM-CGTGGACCTGATGC-MGBNFQ	This study
6	<i>M. africanum</i> West African II	AFRI 1	Rv378404GA	CGGCCGACAGCGAGAA	CCATCACGACCGAATGCTT	6FAM-CTGCAAATCCCGCAGTA-MGBNFQ	VIC-CTGCAAATCCCCACAGT-MGBNFQ	This study

¹Nomenclature according to [2].²Nomenclature according to [20].³Nomenclature according to [72].⁴Position of SNP in reference to the H37Rv genome.⁵6FAM and VIC, fluorescent dyes at the 5'-end of probes; MGBNFQ, MinorGrooveBinder-NonFluorescentQuencher at the 3'-end.

doi:10.1371/journal.pone.0041253.t002

allele (Table S1 and Table 3). Eight data points, for which no allele call was obtained, involved crude lysates, and two were purified DNA samples. For two strains, N1007 and N1032, no allele was called for the *M. bovis*/*M. caprae*-specific SNP, Rv3480645TG, for both crude and purified DNA. To test for a potential deletion in the corresponding genomic region, we designed PCR primers and amplified the region up- and downstream of this SNP. Gel electrophoresis showed an expected amplicon length of approximately 250 bp for H37Rv, but no band for the two strains in question, suggesting that this region was deleted in strains N1007 and N1032 (Figure S1).

For TaqMan PCR, 498 alleles of 504 (98.8%) total data points were successfully called (Table S1 and Table 3). The remaining six data points without allele calls were two crude lysates and four purified DNA samples. There was no case where TaqMan PCR failed to call an allele for both purified and heat-inactivated sample.

Both methods agreed 100% in lineage assignment of the 36 strains comprising the training panel (excluding *M. bovis*, *M. caprae*, *M. microti* and *M. pinnipedi*, as no SNP for *M. bovis*/*M. caprae* was used in the TaqMan assay) (Table 4). Using the prototype spacer definitions [63,64], spoligotyping data confirmed the lineage-assignments of TaqMan and MOL-PCR for 39 of 42 strains (excluding *M. microti*/*M. pinnipedi*). Of the three remaining strains, N0153 and N0051 could not be classified into any of the known spoligotype families. Strain N1024 showed a “Beijing” spoligopattern and was confirmed by TaqMan and MOL-PCR as a Lineage 3 (also known as CAS/Dehli) strain (Table 3). The spoligotype of this strain was described as “Pseudo-Beijing” previously [14].

Next, we used a test panel of 78 clinical isolates from Nepal to validate our methods (Table S2). These samples were mainly crude, heat-inactivated extracts, and no genotyping had been done before. To follow the procedure that would be used as routine SNP-typing, we first performed MOL-PCR, as information for all SNPs is obtained in parallel. Of a total of 624 data points (i.e. 78 strains and eight SNPs), 619 (99.2%) allele calls were obtained (Table S2 and Table 3). The five data points with no calls were from three strains (1671A, 3052B and 3074B). Thereof, two strains (1671A, 3074B) failed to be assigned to a lineage by MOL-PCR

(97.4% sensitivity in lineage assignment). Nevertheless, we obtained ancestral allele calls for seven (1671B) and six (3074B) other SNPs for these strains. We then performed the corresponding TaqMan assay for each sample. TaqMan called alleles for all 78 samples. Again, lineage-assignment was 100% congruent between MOL-PCR and TaqMan.

To evaluate the signal detection limit of MOL-PCR, we used different DNA concentrations as input material. Samples with known DNA concentration were diluted appropriately for different amounts of input DNA. One strain of each lineage was used and MOL-PCR performed with 60 ng, 10 ng, 1 ng, 0.1 ng purified DNA, and 4 µL of crude, heat-inactivated extract. Figures S2, S3, S4, S5, S6, S7 and S8 show a decrease of signal intensities with decreasing amount of DNA. Nevertheless, the use of 0.1 ng purified DNA (CTAB) or 4 µL crude extract still resulted in signals high enough for allelic ratio determination. Signal intensities of the complementary allele were constantly low.

Discussion

Informative SNPs are valuable markers for deep phylogenetic typing of clinical MTBC isolates [11]. Therefore, SNP-typing will likely play an important role for genotype-phenotype association studies in the coming years. Even though routine whole-genome-sequencing, i.e. “genomotyping”, might replace most other typing methods eventually, current limitations such as costs, bioinformatics capacity, the lack of user-friendly software packages, and data storage has been hindering the broad implementation of “genomotyping” [1,11]. Nevertheless, thousands of informative SNPs have become available from whole genome sequences, which can be used for SNP-typing.

In this study, we presented 14 canonical SNPs, comprising two sets of markers for the six main phylogenetic lineages of MTBC, the *M. bovis*/*M. caprae* clade and the “Beijing” sublineage. In contrast to other published SNP sets [23,24], our SNPs were obtained from many whole genome sequences representing the global diversity of MTBC, and thus do not suffer from phylogenetic discovery bias [65]. Phylogenetic discovery bias occurs when SNPs used for subsequent genotyping are initially identified by comparing only a few whole genome sequences (for example when these genomes represent only a subset of all MTBC lineages). As a result, strains belonging to lineages not represented in the initial genome set will falsely appear as intermediates in the corresponding phylogenies [66]. In additional advantage of our SNP sets is that they also identify *Mycobacterium africanum* [67] and discriminate between the West-African I and the West-African II subtypes (i.e. MTBC Lineages 5 and 6). Moreover, the addition of a “Beijing”-specific SNP makes the differentiation between this sublineage and other strains of Lineage 2 (East-Asian) possible.

For the rapid typing of the SNPs described, we have successfully developed a multiplexed assay adapted from MOL-PCR [39]. Our assay uses three oligonucleotides per SNP instead of two, with the LPOs competing for annealing to the template DNA. This allows an allele-call based on allelic ratio rather than signal-to-noise ratio, which increases sensitivity. We further separated the hybridization/ligation from the PCR step, which reduced unspecific signal. We compared the new MOL-PCR assay with TaqMan real-time PCR which was used as standard SNP-typing method in our laboratory. We found an agreement of 100% in classifying strains into the main phylogenetic lineages and an allele-calling sensitivity of 98.6% and 98.8%, respectively, when typing two panels of clinical MTBC strains.

MOL-PCR was sensitive enough to be used with low concentrations of DNA as well as heat-inactivated samples (crude

Table 3. Comparison of MOL-PCR and TaqMan for allele calling and lineage assignment.

		MOL-PCR	TaqMan
Training panel	Total data points	672 (100%) ¹	504 (100%) ¹
	Data points with allele call	662 (98.6%)	498 (98.8%)
	Total number of strains	42 (100%) ²	36 (100%) ³
	Strains assigned to a lineage	42 (100%)	36 (100%)
	Congruence in lineage assignment	100%	
Test panel	Total data points	624 (100%)	78 (100%)
	Data points with allele call	619 (99.2%)	78 (100%)
	Total number of strains	78	78
	Strains assigned to a lineage	76 (97.4%)	78 (100%)
	Congruence in lineage assignment	100%	

¹excluding N/A as indicated in Table S1.

²excluding *M. microti*, *M. pinnipedi*.

³excluding *M. bovis*, *M. caprae*, *M. microti*, *M. pinnipedi*.

doi:10.1371/journal.pone.0041253.t003

Table 4. Lineage assignments of MOL-PCR, TaqMan and spoligotyping for the 46 MTBC strains as in Table S1.

[illegible]

Table 4. Cont.

Lineage assignment				SITVIT lineage ²	SIT ²
MTBC strain	MOL-PCR	TaqMan	Spoligotyping ¹		
N1063	5	5	5	-	-
N1064	5	5	5	AFRI_2	101
N1286	5	5	5	-	-
N1201	6	6	6	-	3476
N1292	6	6	6	-	-
N0091	6	6	6	AFRI_1	181
N0092	6	6	6	AFRI_1	181
N0115	6	6	6	AFRI_1	181
<i>M. bovis</i> BCG Pasteur	<i>M. bovis/M. caprae</i>	(N/A)	<i>M. bovis</i>	BOV_1	482
<i>M. bovis</i> BCG Montreal	<i>M. bovis/M. caprae</i>	(N/A)	<i>M. bovis</i>	BOV_1	482
<i>M. bovis</i> 1290-03	<i>M. bovis/M. caprae</i>	(N/A)	<i>M. bovis</i>	BOV_1	665
<i>M. bovis</i> 8217-02	<i>M. bovis/M. caprae</i>	(N/A)	<i>M. bovis</i>	BOV_1	1037
<i>M. caprae</i> 7140-99	<i>M. bovis/M. caprae</i>	(N/A)	(N/A)	BOV4_CAPRAE	647
<i>M. caprae</i> 8986-99	<i>M. bovis/M. caprae</i>	(N/A)	(N/A)	BOV4_CAPRAE	818
<i>M. microti</i> 1479-00	-	(N/A)	(N/A)	MICROTI	539
<i>M. microti</i> 8753-00	-	(N/A)	(N/A)	PINI1	642
<i>M. pinnipedii</i> 7011-02	-	(N/A)	(N/A)	-	-
<i>M. pinnipedii</i> 7739-01	-	(N/A)	(N/A)	-	-

¹Lineage-classification using spoligotype prototypes as described in [10,63,64].²As described in [64].³No lineage-classification possible based on spoligotyping.⁴"Pseudo-Beijing" as described in [14].

doi:10.1371/journal.pone.0041253.t004

extracts) (Figures S2, S3, S4, S5, S6, S7 and S8). We found that the detection limit of MOL-PCR was as low as 0.1 ng of input DNA. Heat-inactivated samples resulted in similarly high signal levels. This is especially valuable as performing MTBC cultures and DNA extraction is laborious and time-consuming. Even though the signal levels of H₂O control reactions were at time relatively high (Figures S2, S3, S4, S5, S6, S7 and S8), the MFI threshold determination we applied allowed us to control for this, and water samples could not be called false-positive alleles. We recommend to include at least three water samples per 96-well plate, and to use a mean value for the calculation of allelic ratios. Additionally, we always included H₂O control samples for the PCR step to detect potential contamination.

We observed a putative deletion in the region of SNP Rv3480645TG, as we did not obtain signal for neither allele for

two samples (N1007 and N1032). To our knowledge, no genomic deletion has been described to span this SNP [19]. PCR amplification of this region resulted in no visible band for these two samples (Figure S1), which suggested that there might be a deletion present in some Lineage 3 (CAS/Delhi) strains. This highlights the fact that genomic deletions can affect a SNP call, even though the corresponding gene was originally described as essential [54,55]. However, with MOL-PCR we always generated complete allelic information for all other SNPs of the assay. As the canonical SNPs defining different MTBC lineages are mutually exclusive (except Lineage 2 and Beijing), the lack of an allelic call for one SNP can be neglected if the alleles at other loci are called successfully.

A limitation in the development of the MOL-PCR assay was the low efficiency of design and validation of oligonucleotides. For

each lineage, we had to design up to 5 sets of probes targeting different SNPs, because many had to be rejected during validation, most often because of unfavorable allelic or signal-to-noise ratios or lack of signal. The use of long (i.e. 40–140 bp) oligonucleotides makes hairpin and homodimer formations likely, and multiplexing probes adds a level of complexity due to a high combinatorial number of possible heterodimer formations. We used different software for homo- and heterodimer analysis, but still were not able to fully predict the success of probes. Recently, an online tool for MOL-PCR specific oligonucleotide design and evaluation became available (moligodesigner.lanl.gov) [58], which could strongly facilitate the design of such oligonucleotides. Furthermore, an oligonucleotide design tool for copy-number-variation analysis of human, mouse and rat genomes with MLPA and MOL-PCR has been described [68,69], and might be extended to SNP analysis and bacterial genomes.

Four other methods have been described for multiplexed SNP-typing on the Luminex platform, namely Direct Hybridization, Allele-Specific Primer Extension, Single-Base-Extension and Oligonucleotide Ligation [40,41]. In contrast to MOL-PCR, these methods use multiplex-PCR for amplification of template DNA. This makes high multiplex-levels difficult, as a multiplex PCR is usually the limiting factor in multiplexed assays. In our assay, 24 oligonucleotides of 48 to 80 bp lengths were included for eight SNPs. A number of additional SNPs could be interrogated in the same reaction, i.e. for sublineages, drug resistance mutations [25] or outbreak-specific SNPs [70].

In addition to MOL-PCR, we also developed TaqMan SNP-assays for robust and rapid singleplex-typing of all main MTBC lineages. TaqMan SNP-typing is a well-established method for genotyping and available commercially. This facilitates the design and validation of probes, and provides standardized reaction conditions. Furthermore, criteria of TaqMan primer and probe design are less strict than in MOL-PCR. TaqMan makes use of quantitative PCR, which renders the method particularly sensitive. On the other hand, multiplexing of TaqMan assays is demanding and the degree of multiplexing is limited. Hence, we recommend using MOL-PCR for strain collections with no *a priori* information. TaqMan SNP-typing, on the other hand, is most suitable for confirmation of MOL-PCR results, or for strain collections for which previous information about genotypes is available.

Costs for reagents and consumables are another important consideration when implementing a new technique. Similar to other methods, running costs of MOL-PCR per SNP decrease with increasing number of multiplexed SNPs. We have estimated the minimal reagent costs per SNP to be ~0.8 Euro when run in a singleplex reaction. Reagent costs drop to ~0.15 Euro per SNP when running an 8-plex reaction. The cost of one TaqMan reaction was estimated at ~0.9 Euro per SNP. For both assays, the initial setup costs are relatively high due to purchase of customized and labeled oligonucleotides. Furthermore, both methods require an expensive piece of equipment in form of a Luminex or a real-time-PCR machine. However, both of these devices can be used for a wide variety of applications in addition to SNP-typing (e.g. spoligotyping on the Luminex platform [71]), and many laboratories already have access to such equipment.

In conclusion, we propose a new set of canonical SNPs specific for the main phylogenetic lineages of MTBC. When combined with multiplexed MOL-PCR or singleplex TaqMan SNP-typing, these SNPs are ideal for phylotyping of strain collections from a small to a large scale. These assays provide a new basis for phylogenetically robust classification of clinical isolates.

Supporting Information

Figure S1 Agarose gel of PCR product of Rv3114 covering the SNP specific for *M. bovis*/*M. caprae* (Rv3480645TG). H37Rv was used as positive control, whereas N1007 and N1032 were the samples that did not result in any signal in MOL-PCR. Ladder is Hyperladder II (Bioline). (PDF)

Figure S2 Dilution series of DNA from a clinical isolate of Lineage 1. Strain N1004 was previously characterized as strain of Lineage 1. Median reporter fluorescence intensities (Alexa532-primer) of MOL-PCR are shown for each bead region and by different DNA amounts (CTAB-extracted) or crude extract (heat-inactivated). Error bars represent standard deviations calculated from triplicates. (PDF)

Figure S3 Dilution series of DNA from a clinical isolate of Lineage 2/Beijing. Strain N1053 was previously characterized as strain of Lineage 2/Beijing. Median reporter fluorescence intensities (Alexa532-primer) of MOL-PCR are shown for each bead region and by different DNA amounts (CTAB-extracted) or crude extract (heat-inactivated). Error bars represent standard deviations calculated from triplicates. (PDF)

Figure S4 Dilution series of DNA from a clinical isolate of Lineage 3. Strain N1144 was previously characterized as strain of Lineage 3. Median reporter fluorescence intensities (Alexa532-primer) of MOL-PCR are shown for each bead region and by different DNA amounts (CTAB-extracted) or crude extract (heat-inactivated). Error bars represent standard deviations calculated from triplicates. (PDF)

Figure S5 Dilution series of DNA from a clinical isolate of Lineage 4. Strain N1015 was previously characterized as strain of Lineage 4. Median reporter fluorescence intensities (Alexa532-primer) of MOL-PCR are shown for each bead region and by different DNA amounts (CTAB-extracted) or crude extract (heat-inactivated). Error bars represent standard deviations calculated from triplicates. (PDF)

Figure S6 Dilution series of DNA from a clinical isolate of Lineage 5. Strain N1286 was previously characterized as strain of Lineage 5. Median reporter fluorescence intensities (Alexa532-primer) of MOL-PCR are shown for each bead region and by different DNA amounts (CTAB-extracted) or crude extract (heat-inactivated). Error bars represent standard deviations calculated from triplicates. (PDF)

Figure S7 Dilution series of DNA from a clinical isolate of Lineage 6. Strain N1200 was previously characterized as strain of Lineage 6. Median reporter fluorescence intensities (Alexa532-primer) of MOL-PCR are shown for each bead region and by different DNA amounts (CTAB-extracted) or crude extract (heat-inactivated). Error bars represent standard deviations calculated from triplicates. (PDF)

Figure S8 Dilution series of DNA from *M. bovis* BCG Pasteur. Median reporter fluorescence intensities (Alexa532-primer) of MOL-PCR are shown for each bead region and by different DNA amounts (CTAB-extracted) or crude extract (heat-

inactivated). Error bars represent standard deviations calculated from triplicates.
(PDF)

Table S1 MOL-PCR and TaqMan allele calls for the training panel of 46 well-characterized MTBC strains.
(XLS)

Table S2 MOL-PCR and TaqMan allele calls for the test panel of 78 MTBC strains with no previous genotyping information.
(XLS)

References

- Kato-Maeda M, Metcalfe JZ, Flores L (2011) Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. *Future Microbiol* 6: 203–216. doi:10.2217/fmb.10.165.
- Coscolla M, Gagneux S (2010) Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov Today Dis Mech* 7: e43–e59. doi:10.1016/j.ddmec.2010.09.004.
- Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NTN, et al. (2008) The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* 4: e1000034. doi:10.1371/journal.ppat.1000034.
- Nicol MP, Wilkinson RJ (2008) The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. *Trans R Soc Trop Med Hyg* 102: 955–965. doi:10.1016/j.trstmh.2008.03.025.
- Thwaites G, Caws M, Chau TTH, D'Sa A, Lan NTN, et al. (2008) Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J Clin Microbiol* 46: 1363–1368. doi:10.1128/JCM.02180-07.
- de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, et al. (2008) Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *J Infect Dis* 198: 1037–1043. doi:10.1086/591504.
- Homolka S, Niemann S, Russell DG, Rohde KH (2010) Functional genetic diversity among *Mycobacterium tuberculosis* complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. *PLoS Pathog* 6: e1000988. doi:10.1371/journal.ppat.1000988.
- Portevin D, Gagneux S, Comas I, Young D (2011) Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog* 7: e1001307. doi:10.1371/journal.ppat.1001307.
- Click ES, Moonan PK, Winston CA, Cowan LS, Oeltmann JE (2012) Relationship between *Mycobacterium tuberculosis* phylogenetic lineage and clinical site of tuberculosis. *Clin Infect Dis* 54: 211–219. doi:10.1093/cid/cir788.
- Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE* 4: e7815. doi:10.1371/journal.pone.0007815.
- Schürch AC, van Soolingen D (2012) DNA fingerprinting of *Mycobacterium tuberculosis*: From phage typing to whole-genome sequencing. *Infect Genet Evol* 12: 602–609. doi:10.1016/j.meegid.2011.08.032.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 35: 907–914.
- Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, et al. (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 44: 4498–4510. doi:10.1128/JCM.01392-06.
- Fenner L, Malla B, Ninet B, Dubuis O, Stucki D, et al. (2011) “Pseudo-Beijing”: Evidence for convergent evolution in the direct repeat region of *Mycobacterium tuberculosis*. *PLoS ONE* 6: e24737. doi:10.1371/journal.pone.0024737.
- Flores L, Van T, Narayanan S, DeRiemer K, Kato-Maeda M, et al. (2007) Large sequence polymorphisms classify *Mycobacterium tuberculosis* strains with ancestral spoligotyping patterns. *J Clin Microbiol* 45: 3393–3395. doi:10.1128/JCM.00828-07.
- McEvoy CRE, Falmer AA, van Pittius NCG, Victor TC, van Helden PD, et al. (2007) The role of IS6110 in the evolution of *Mycobacterium tuberculosis*. *Tuberculosis* 87: 393–404. doi:10.1016/j.tube.2007.05.010.
- Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD (1993) Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* 10: 1057–1065.
- Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, et al. (2000) Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* 36: 762–771.
- Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, et al. (2004) Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci USA* 101: 4865–4870. doi:10.1073/pnas.0305634101.
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, et al. (2006) Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 103: 2869–2873. doi:10.1073/pnas.0511240103.
- Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA* 94: 9869–9874.
- Baker L, Brown T, Maiden MC, Drobniewski F (2004) Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerging Infect Dis* 10: 1568–1577.
- Filioli I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, et al. (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol* 188: 759–772. doi:10.1128/JB.188.2.759-772.2006.
- Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, et al. (2006) Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J Infect Dis* 193: 121–128. doi:10.1086/498574.
- Bergval IL, Vijzelaar RNCP, Dalla Costa ER, Schuitema ARJ, Oskam L, et al. (2008) Development of multiplex assay for rapid characterization of *Mycobacterium tuberculosis*. *J Clin Microbiol* 46: 689–699. doi:10.1128/JCM.01821-07.
- Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, et al. (2008) High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 6: e311. doi:10.1371/journal.pbio.0060311.
- Abadia E, Zhang J, dos Vultos T, Ritacco V, Kremer K, et al. (2010) Resolving lineage assignment on *Mycobacterium tuberculosis* clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. *Infect Genet Evol* 10: 1066–1074. doi:10.1016/j.meegid.2010.07.006.
- Bouakaze C, Keyser C, de Martino SJ, Sougakoff W, Veziris N, et al. (2010) Identification and genotyping of *Mycobacterium tuberculosis* complex species by use of a SNaPshot Minisequencing-based assay. *J Clin Microbiol* 48: 1758–1766. doi:10.1128/JCM.02255-09.
- Bouakaze C, Keyser C, Gonzalez A, Sougakoff W, Veziris N, et al. (2011) Matrix-assisted laser desorption/ionization-time of flight mass spectrometry-based single nucleotide polymorphism genotyping assay using iPLEX gold technology for identification of *Mycobacterium tuberculosis* complex species and lineages. *J Clin Microbiol* 49: 3292–3299. doi:10.1128/JCM.00744-11.
- Kim S, Misra A (2007) SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng* 9: 289–320. doi:10.1146/annurev.bioeng.9.060906.152037.
- Wang L, Luhm R, Lei M (2007) SNP and mutation analysis. *Adv Exp Med Biol* 593: 105–116. doi:10.1007/978-0-387-39978-2_11.
- Espinosa de los Monteros LE, Galán JC, Gutiérrez M, Samper S, García Marín JF, et al. (1998) Allele-specific PCR method based on *pncA* and *oxyR* sequences for distinguishing *Mycobacterium bovis* from *Mycobacterium tuberculosis*: intraspecific *M. bovis* *pncA* sequence polymorphism. *J Clin Microbiol* 36: 239–242.
- Halse TA, Escuyer VE, Musser KA (2011) Evaluation of a single-tube multiplex real-time PCR for differentiation of members of the *Mycobacterium tuberculosis* complex in clinical specimens. *J Clin Microbiol* 49: 2562–2567. doi:10.1128/JCM.00467-11.
- Mestre O, Luo T, Dos Vultos T, Kremer K, Murray A, et al. (2011) Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS ONE* 6: e16020. doi:10.1371/journal.pone.0016020.
- Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, et al. (2010) Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* 42: 498–503. doi:10.1038/ng.590.
- Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, et al. (2004) Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infection, Genetics and Evolution* 4: 205–213. doi:10.1016/j.meegid.2004.02.005.
- Caws M, Thwaites G, Stepniewska K, Nguyen TNL, Nguyen THD, et al. (2006) Beijing genotype of *Mycobacterium tuberculosis* is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of

Acknowledgments

We would like to thank Bhawana Shrestha, Director of the German Nepal Tuberculosis Project, Kathmandu, Nepal, for help with sample collection, Alina Deshpande for helpful technical advice, and Jean Pieters and Stefan Niemann for providing MTBC strains.

Author Contributions

Conceived and designed the experiments: DS BM LF IC SG. Performed the experiments: DS BM SH TH JF SB IC MC. Analyzed the data: DS BM IC MC SG. Contributed reagents/materials/analysis tools: BM DY SG. Wrote the paper: DS BM SB LF IC SG.

- tuberculous meningitis. *J Clin Microbiol* 44: 3934–3939. doi:10.1128/JCM.01181-06.
38. Parvati I, van Crevel R, van Soolingen D (2010) Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect Dis* 10: 103–111. doi:10.1016/S1473-3099(09)70330-5.
39. Deshpande A, Gans J, Graves SW, Green L, Taylor L, et al. (2010) A rapid multiplex assay for nucleic acid-based diagnostics. *J Microbiol Methods* 80: 155–163. doi:10.1016/j.jmimet.2009.12.001.
40. Dunbar SA (2006) Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clin Chim Acta* 363: 71–82. doi:10.1016/j.cccn.2005.06.023.
41. Lee S-H, Walker DR, Cregan PB, Boerma HR (2004) Comparison of four flow cytometric SNP detection assays and their use in plant improvement. *Theor Appl Genet* 110: 167–174. doi:10.1007/s00122-004-1827-1.
42. Clotilde LM, Bernard C 4th, Hartman GL, Lau DK, Carter JM (2011) Microbead-based immunoassay for simultaneous detection of Shiga toxins and isolation of *Escherichia coli* O157 in foods. *J Food Prot* 74: 373–379. doi:10.4315/0362-028X.JFP-10-344.
43. Yesilkaya H, Meacci F, Niemann S, Hillemann D, Rüscher-Gerdes S, et al. (2006) Evaluation of molecular-beacon, TaqMan, and fluorescence resonance energy transfer probes for detection of antibiotic resistance-conferring single nucleotide polymorphisms in mixed *Mycobacterium tuberculosis* DNA extracts. *J Clin Microbiol* 44: 3826–3829. doi:10.1128/JCM.00225-06.
44. Ben Shabat M, Mikula I, Gerchman I, Lysnyansky I (2010) Development and evaluation of a novel single-nucleotide-polymorphism real-time PCR assay for rapid detection of fluoroquinolone-resistant *Mycoplasma bovis*. *J Clin Microbiol* 48: 2909–2915. doi:10.1128/JCM.00699-10.
45. Milner DA Jr, Varetta J, Valim C, Montgomery J, Daniels RF, et al. (2012) Human cerebral malaria and *Plasmodium falciparum* genotypes in Malawi. *Malaria Journal* 11: 35. doi:10.1186/1475-2875-11-35.
46. Nübel U, Nitsche A, Leyer F, Strommenger B, Witte W (2012) Single-nucleotide polymorphism genotyping identifies a locally endemic clone of methicillin-resistant *Staphylococcus aureus*. *PLoS ONE* 7: e32698. doi:10.1371/journal.pone.0032698.
47. Allix-Béguec C, Harmsen D, Weniger T, Supply P, Niemann S (2008) Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *J Clin Microbiol* 46: 2692–2699. doi:10.1128/JCM.00540-08.
48. Yeboah-Manu D, Asante-Poku A, Bodmer T, Stucki D, Koram K, et al. (2011) Genotypic diversity and drug susceptibility patterns among *M. tuberculosis* complex isolates from South-Western Ghana. *PLoS ONE* 6: e21906. doi:10.1371/journal.pone.0021906.
49. Fenner L, Gagneux S, Helbling P, Battegay M, Rieder HL, et al. (2012) *Mycobacterium tuberculosis* transmission in a country with low tuberculosis incidence: role of immigration and HIV infection. *J Clin Microbiol* 50: 388–395. doi:10.1128/JCM.05392-11.
50. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 31: 406–409.
51. Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, et al. (2012) The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. *PLoS Negl Trop Dis* 6: e1552. doi:10.1371/journal.pntd.0001552.
52. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739. doi:10.1093/molbev/msr121.
53. Maddison WP, Maddison DR (2011) Mesquite: a modular system for evolutionary analysis. Version 2.73. Available: <http://mesquiteproject.org/>.
54. Sasseti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci USA* 100: 12989–12994. doi:10.1073/pnas.2134250100.
55. Sasseti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48: 77–84.
56. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537–544. doi:10.1038/31159.
57. Schouten JP, McElgunn CJ, Waaij R, Zwijsenburg D, Diepvens F, et al. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 30: e57.
58. Song J, Li P-E, Gans J, Vuyisich M, Deshpande A, et al. (2010) Simultaneous pathogen detection and antibiotic resistance characterization using SNP-based multiplexed oligonucleotide ligation-PCR (MOL-PCR). *Adv Exp Med Biol* 680: 455–464. doi:10.1007/978-1-4419-5913-3_51.
59. Okonechnikov K, Golosova O, Fursov M (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics (Oxford, England)* 28: 1166–1167. doi:10.1093/bioinformatics/bts091.
60. Kaderali L, Deshpande A, Nolan JP, White PS (2003) Primer design for multiplexed genotyping. *Nucleic Acids Res* 31: 1796–1802. doi:10.1093/nar/kgk267.
61. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *BioTechniques* 37: 226–231.
62. Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5'–3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci USA* 88: 7276–7280.
63. Kato-Maeda M, Gagneux S, Flores LL, Kim EY, Small PM, et al. (2011) Strain classification of *Mycobacterium tuberculosis*: congruence between large sequence polymorphisms and spoligotypes. *Int J Tuberc Lung Dis* 15: 131–133.
64. Demay C, Liens B, Burguière T, Hill V, Couvin D, et al. (2012) SITVITWEB – A publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 12: 755–766. doi:10.1016/j.meegid.2012.02.004.
65. Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, et al. (2004) Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci USA* 101: 13536–13541. doi:10.1073/pnas.0403844101.
66. Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009) Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 7: 537–544. doi:10.1038/nrmicro2165.
67. de Jong BC, Antonio M, Gagneux S (2010) *Mycobacterium africanum*—review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis* 4: e744. doi:10.1371/journal.pntd.0000744.
68. Zhi J, Hatchwell E (2008) Human MLPA Probe Design (H-MAPD): a probe design tool for both electrophoresis-based and bead-coupled human multiplex ligation-dependent probe amplification assays. *BMC Genomics* 9: 407. doi:10.1186/1471-2164-9-407.
69. Zhi J (2010) MAPD: a probe design suite for multiplex ligation-dependent probe amplification assays. *BMC Research Notes* 3: 137. doi:10.1186/1756-0500-3-137.
70. Schürch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, et al. (2010) High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol* 48: 3403–3406. doi:10.1128/JCM.00370-10.
71. Cowan LS, Diem L, Brake MC, Crawford JT (2004) Transfer of a *Mycobacterium tuberculosis* genotyping method, Spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. *J Clin Microbiol* 42: 474–477.
72. Filliol I, Driscoll JR, van Soolingen D, Kreiswirth BN, Kremer K, et al. (2003) Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J Clin Microbiol* 41: 1963–1970.